Dunn S, Wilkinson S, Ford A.

Spatial Structure and Evolution of Infrastructure Networks.

***Sustainable Cities and Society* 2016, 27, 23-31.**

DOI link to article:

http://dx.doi.org/10.1016/j.scs.2016.08.011

Date deposited:

04/10/2016

Sustainable Cities and Society 27 (2016) 23–31

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Spatial structure and evolution of infrastructure networks

Sarah Dunn [a,*], Sean Wilkinson [b], Alistair Ford [c]

[a] Lecturer in Structural Engineering, School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[b] Senior Lecturer in Structural Engineering, School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[c] Researcher in Geomatics, School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

ARTICLE INFO

Article history:
Received 11 February 2016
Received in revised form 2 August 2016
Accepted 4 August 2016
Available online 9 August 2016

Keywords:
Network theory
Spatial networks
Transportation networks
Air traffic networks
Network growth
Infrastructure
Resilience

ABSTRACT

Critical infrastructure systems are essential components of a nation's assets. They support economic development, enable growth and help to protect against hazard. In recent years, the demands placed upon these systems have been rapidly increasing, partly due to the shift from rural to urban living and partly due to increasing wealth. For example, in 1960, only 34% of the world's population lived in cities whereas, in 2014, this figure had risen rapidly to 54%. This shift has also caused a massive explosion in urban infrastructure systems and therefore a proportionately greater risk to social cohesion through the potential loss of critical infrastructure due to natural or manmade hazard. While it is possible to model the performance of these systems, the complexity of them makes it difficult to assess their contribution to economic development or their resilience to hazard. This deficiency stems from our inability to identify key generic features that would enable us to simplify the task and hence conduct probabilistic assessments or to recognise the underlying drivers that govern their evolution and thus enable us to make robust future decisions. In this paper we present an algorithm that can generate spatial nodal layouts which share a number of non-trivial features common to several types of real world networks. The synthetic networks generated by the algorithm can be used in planning studies to see how infrastructure may evolve in the future, considering alternate planning or policy scenarios for example, or in other scenario based assessments, such as hazard tolerance studies.

1. Introduction

In our modern society, cities are one of the main the key factors in the development of a country's social prosperity and economic growth. They bring together people, jobs, facilities and all the other inputs necessary to generate wealth and improve living standards (Florida 2011). Over 54% of the world's population currently live in cities and it is anticipated that, on average, cities across Asia will see a 91% growth in their population over the next decade (Knight Frank 2015). Cities can also enable risk mitigation and increase community resilience to hazard events (Lizarralde, Chmutina, Bosher, & Dainty, 2015; Miller 2015), promote the sustainable growth of a country (Ruza, Kim, Leung, Kam, & Ng, 2015; Sharifi 2016) and potentially help to mitigate the impacts of climate change (Spataru, Drummond, Zafeiratou, & Barrett, 2015; Kwan & Hashim, 2016). Within our cities, it could be argued that the most important elements are the critical infrastructure systems. They facilitate travel, support living standards (through the provision of clean water and electrical supplies) and provide a means of communication, amongst others. One approach to the analysis of these infrastructure systems has been through the application of complex networks, or network graph theory, models (Dunn, Fu, Wilkinson, & Dawson, 2013; Wilkinson, Dunn, & Ma, 2011).

The study of complex networks has developed rapidly in the past few years and is seemingly driven by the desire to understand the fundamental properties that are generic to many of these. In this effort, the primary research focus has been on understanding why the connections between components establish themselves resulting in complex systems with specific architectures (i.e. network classes) (Gorman & Kulkarni, 2004), a view to understanding the resilience of these systems (Albert, Jeong, & Barabasi, 2000; Lordan, Sallan, & Simo, 2014; Zhang, Miller-Hooks, & Denny, 2015). There have been a number of network classes established in the literature, starting with the basic random graph model (Erdos & Renyi, 1960) and moving through to the more sophisticated network models, such as small-world networks (Watts & Strogatz 1998) and more recently scale-free (Barabasi & Albert, 1999) and exponential networks (Liu & Tang 2005), a brief description of these network classes is provided by Dunn et al. (Dunn et al., 2013). The latter two models have both been shown to be good representations of

* Corresponding author.
E-mail address: sarah.dunn@ncl.ac.uk (S. Dunn).

http://dx.doi.org/10.1016/j.scs.2016.08.011
2210-6707/

many real world networks (Albert, Jeong, & Barabasi, 1999; Crucitti, Latora, & Marchiori, 2004); however, in all of this research, spatial influences have not been considered. As the study of real world networks moves from ethereal networks, such as social networks or the World-Wide Web, and networks that, although tangible, we do not really physically interact with (such as the Internet), to networks where physically interactions are an important factor (e.g. a transportation network) space suddenly becomes an important factor. For these systems there has been very little research on how space may influence their development and structure. The little work that has studied real world spatial networks still focuses mainly on characterising the topology of the system (into one of the network classes) (Carvalho et al., 2009; Sen et al., 2003; Sole, Rosas-Casals, Corominas-Murtra, & Valverde, 2008), while the spatial element of the same network receives less attention – if not neglected entirely (Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006). This spatial component may not seem important, as it could be argued that it is the arrangement of links which defines the characteristics of the network; however, as shown by Wilkinson et al. (2011) in their study of the disruption caused to the European air traffic network (EATN) by the Eyjafjallajökull volcanic event in 2010, space can not only have a significant influence on the layout of the network, but is actually crucial in defining its tolerance to natural hazards. Other studies have considered the scale and shape of cities, by studying real world examples and attempting to replicate their structure in order to understand their drivers for growth, including Bettencourt, Lobo, Helbing, Kuhnert, and West (2007), Bettencourt, Lobo, and West (2008), Batty (2008) and Batty (2013). However, the current understanding of the structure of a city "*remains limited*" (Bettencourt 2013) and the models developed in these studies are complex, requiring detailed inputs in order to function.

The vast majority of network theory studies ignore the 'rules' governing the location of nodes within spatial networks, in favour of the 'rules' governing the formation of links (Gastner & Newman, 2006a,b). For example, in their paper Wilkinson et al. (2011) developed an algorithm to generate synthetic networks for the EATN; however their study was deficient as they only considered the global spatial structure of the nodes in the synthetic networks (making the simplifying assumption that they were distributed uniformly with distance) and therefore the smaller scale features of the network (e.g. the clustering of real airports) were not captured. Other studies have focused on the optimal design of spatial distribution networks (Gastner & Newman, 2006a,b), attempting to minimise the mean distance between a member of the population and the nearest facility (e.g. hospitals). Whilst, network theory has ignored this spatial structure, other studies have focused primarily on this problem, with some success. One of the most notable techniques is cellular automata, which has been used to predict urban growth around cities, including: San Francisco (Clarke, Hoppen, & Gaydos, 1995), Washington/Baltimore (Clarke & Gaydos, 1998) and Guangzhou (China) (Wu 2002). The 'rules' used in these models govern the location of nodes within the study area as the model grows over a given timeframe. The models require the initial input of a number of layers of data describing the initial conditions in the study area, which are updated as the model runs: (1) digital elevation of the study area, (2) the location of the initial settlements, (3) historical transportation layers (e.g. road network) and (4) a layer showing excluded areas (e.g. national parks, water bodies, etc.). This data is gathered from historical maps, air photos and digital maps; and as the data is obtained from a variety of sources there are often problems with assembling the dataset, including: inconsistent dimensions of features, generalisation in historical maps, different projections of the study area and different coordinate systems. As such the main disadvantage in these studies is that the accuracy of the results is highly dependent on the size of the input dataset and on the quality and quantity of the historical data (Clarke & Gaydos, 1998), indeed unless sufficient historical data for the study area is obtained then it is not possible to generate a synthetic model.

In this paper we propose a new method to generate a spatial network, which can be combined with traditional graph theory network generation algorithms to develop a fully synthetic spatial network which shares several non-trivial features of real world spatial networks. The starting point for this model is to simplify the ideas behind cellular automata to generate synthetic nodal layouts for real world networks, as well as more generic nodal layouts (which can be used in tests for network resilience, for example). Specifically, we incorporate a set of initial conditions, which are allowed to evolve over time as the network 'grows' (or 'expands'); however, unlike previous cellular automata, this algorithm uses only one dataset for the study area to provide the initial inputs, rather than using many potentially inaccurate historical datasets. We use three real world infrastructure networks to derive the relationships used in the algorithm, terming these the "development datasets". We then validate the algorithm by using it to generate proxies for two different real world networks, which have not been used in the development of the algorithm, termed the "validation datasets". In these two validation networks, we consider both the 'rules' governing the location of nodes and those governing the formation of links. However, in this study we do not claim that our work describes the complete 'rules' that govern spatial networks, as this is a very complex problem, but our work forms an important first step in characterising the underlying resilience of geographically distributed networks to spatial hazard that is missing from purely topological studies. We demonstrate how our model can be used in future research to consider how different drivers (e.g. population density, city location or policy) can impact upon the resulting spatial characteristics of real world systems.

2. Characteristics of real world networks

In order to develop an algorithm which can generate proxies for real world nodal layouts, the characteristics of these real world networks need to be established and incorporated into the algorithm. To achieve this, we use three "development datasets" to investigate the location of US airports (682 airports), UK rail stations (2605 stations) and 33 kV electricity substations over an area of the UK (526 substations). We use these datasets to represent a range of different infrastructures over different geographical areas. Fig. 1 shows the spatial layout of these three datasets along with their associated spatial distributions.

From Fig. 1(a, c, e) it can be seen that the three datasets are visually very different. The UK rail network appears to have much denser clusters of nodes than those of US airports and 33 kV electricity substations. The spatial distributions for these nodal layouts (Fig. 1(b, d, f)) also show that there are differences between the three datasets. The UK rail stations form distinct bi-linear distributions due to the area of high nodal density located around London, which is also close to the location of the geographic centre. Whereas the US airports and 33 kV substations form a more linear spatial distribution. From the layouts of these two datasets it can be seen that they have a much more uniform spread of nodes over the whole geographic area, although individual clusters of nodes can still be distinguished. It is worth noting, that in this paper, we use the term 'cluster' to define a visually identifiable group of nodes (i.e. a spatial area of particularly high nodal density).

In order to generate proxies for real world networks, the algorithm must be able to not only generate these large scale distributions, but also replicate the smaller scale features of the individual clusters. Additionally, as they are dynamic networks (i.e.

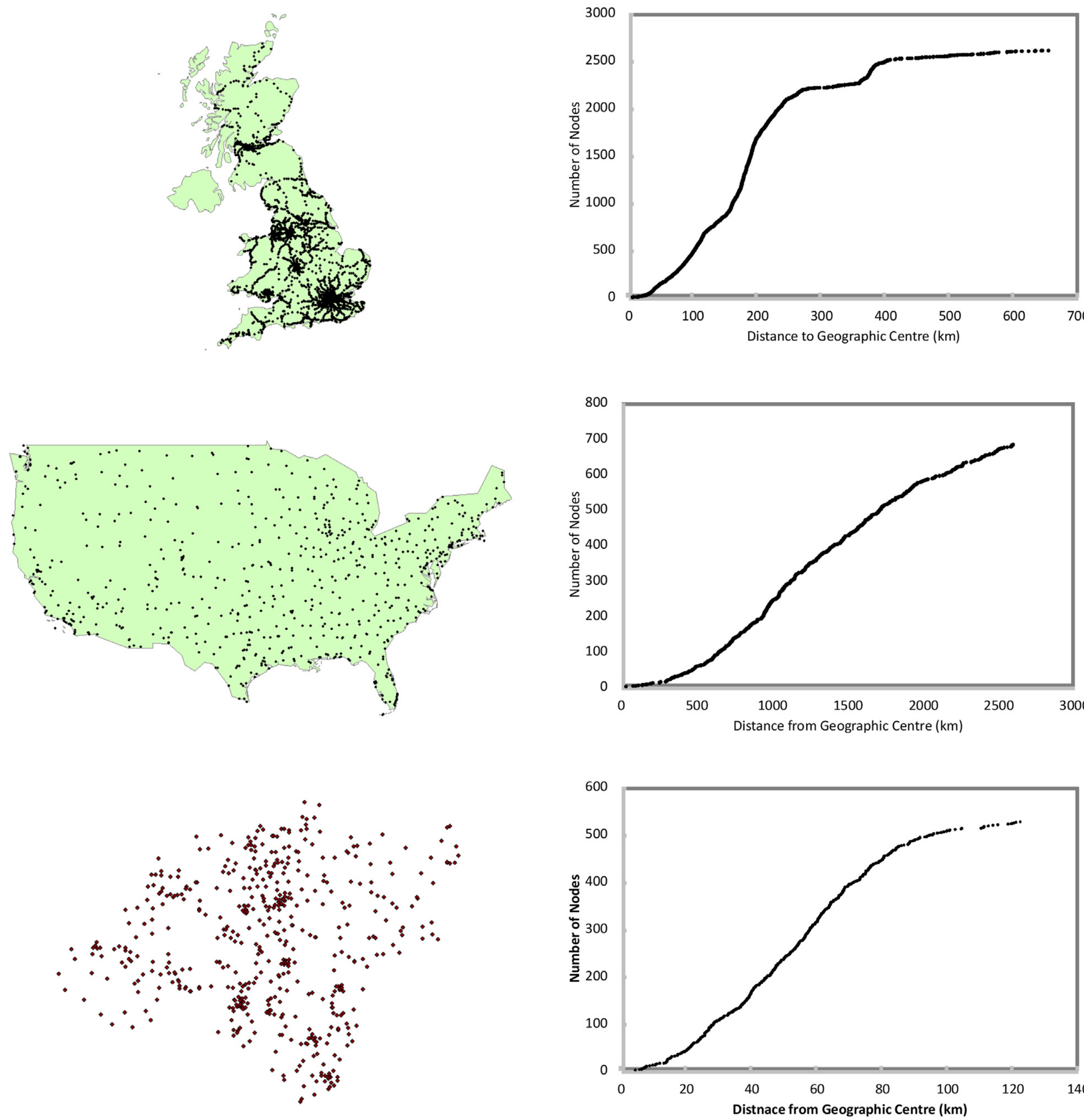

Fig. 1. Showing (a) the spatial layout of nodes for the UK railway dataset (obtained from (data.gov.uk, 2015)) and (b) its associated spatial distribution, (c) the US air traffic network (obtained from (Openflights, 2010)) and (d) its spatial distribution and finally (e) a subset of UK 33 kV substations and (f) its spatial distribution. The spatial distributions were obtained by first calculating the geographic centre of the network and then plotting the number of nodes within a given radius (for example, the geographic centre for the UK rail network is located approximately 75 km north of Birmingham).

Table 1
Showing a summer of the input parameters used in the algorithm to generate the proxy nodal layouts. A more detailed explanation of these parameters is provided in Section 3 and their application to real-world problems is discussed in Section 4.

Input Parameter	Description
Seed Nodes	Defines the location of the starting settlements in the case study area.
Initial Radius (of Seed Nodes)	Defines the “type” of settlement individual seed nodes represent (the higher the value the more dense the settlement, e.g. city).
Cluster Density (C_D)	Determines the density of the global infrastructure in the case study area (the higher the value the more dispersed the infrastructure, e.g. airports).
Proportion of Nodes allowed to form outside the influence of a cluster	Used to simulate a rural environment over the whole of the case study area, by allowing infrastructure to develop outside the “seed node clusters” (the higher the value the more infrastructure forms outside the seed node cluster influence).

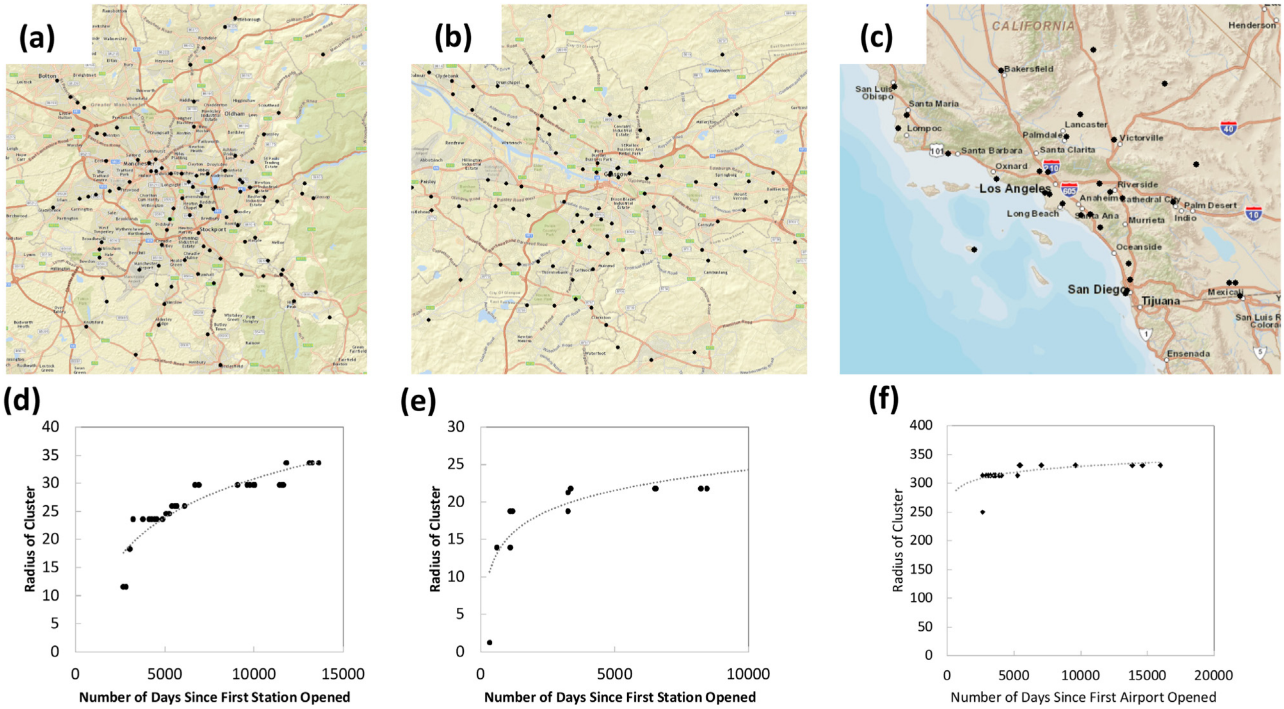

Fig. 2. Showing an individual clusters from (a) UK railway stations (Manchester), (b) UK railway stations (Glasgow) and (c) US airports (Los Angeles). Also showing how the radius (defined as the distance from the oldest node in the cluster to the furthest node) of each cluster changes with time have been plotted for each cluster: (d) Manchester railway stations, (e) Glasgow railway stations and (f) Los Angeles airports.

they expand and grow over time), the growth of these individual clusters must also be modelled. Clusters of nodes from all three datasets were identified and assessed, however, due to space limitations we present two clusters from the UK rail dataset and one from the US air traffic network dataset in this paper. These clusters have been isolated through the use of Kernel Density Images generated using ArcGIS software and are shown in Fig. 2, where it can be seen that each cluster is approximately circular in shape. To quantify the 'growth' of each cluster, we first establish the oldest node in each cluster (i.e. the first railway station opened) and set this as the cluster midpoint. We then introduce nodes to the cluster in the order that they were opened and define the radius of the

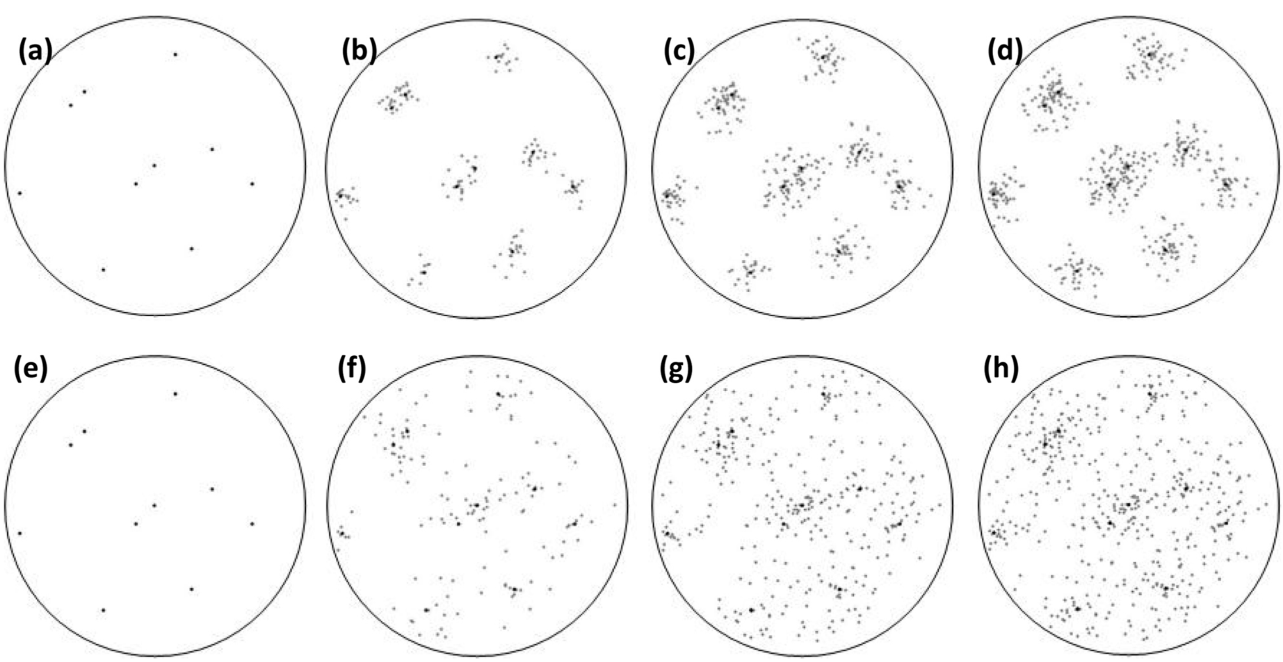

Fig. 3. Showing the progression of the clustering algorithm for two generated networks ((a)–(d) and (e)–(h)) with different C_D values, 200 and 400 respectively. It is worth noting that these C_D values are relative and dimensionless. Where the black dots represent the starting nodes and the grey dots show the added nodes, the outer circle defines the spatial boundary of the network. (a) and (e) show the seed nodes (all of the starting nodes have the same radius); (b) and (f) show the layout after 150 nodes have been added; (c) and (g) show the layout after 350 nodes have been added; (d) and (h) show the final nodal layout. It can be seen that the larger C_D value results in a nodal layout that has visually less dense clusters than that of the smaller C_D value.

Table 2
Showing the input values used to generate the proxy nodal layouts.

Dataset	UK Rail Stations	US Airports	33 kV Substations
Proportion of Seed Nodes	<1%	~1%	~2%
Starting radii used	40, 80, 100	50, 60, 80	10, 50
C_D (global density)	4	33	2
Proportion of Nodes allowed to form outside the influence of a cluster	20%	80%	70%

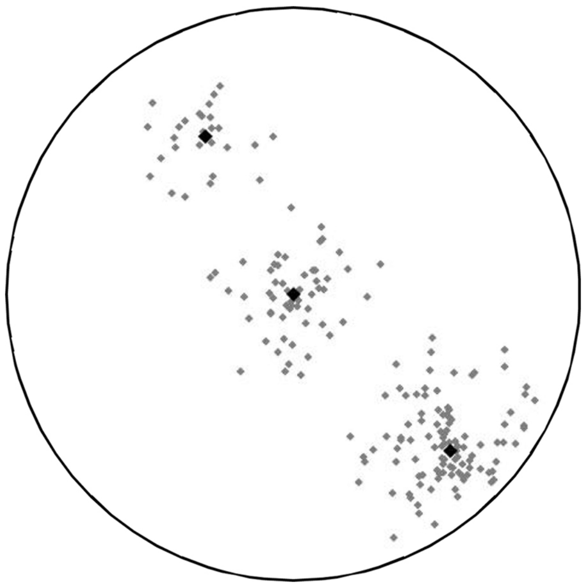

Fig. 4. Showing three seed nodes (black) with different radii values and the subsequent 200 added nodes (grey). The top left starting node has a radius value of 800 and forms the least dense cluster, the bottom right node has a radius value of 10 and forms the densest cluster, whilst the central node has a radius value between these two extremes (500) and therefore has a density between the other two starting nodes. Again it is worth noting that in a similar manner to the C_D value, these values are relative and dimensionless.

cluster as the distance from the cluster midpoint to the furthest store or station in the cluster. This data is plotted in Fig. 2 (d, e, f) against the number of days since the first rail station opened. From this figure, it can be seen that for both clusters the first few nodes added to each cluster increase the radius rapidly, but then only increases slightly, if at all, with further opened nodes. This is due to the opening of one node close to the outer boundary of the cluster in an early time-step. The remaining nodes then open within this boundary, meaning that the radius of the cluster does not change, or changes only marginally, after a few nodes have been created.

From the analysis of these three real world datasets it can be concluded that, even over the same study area, real world nodal layouts can show quite different characteristics. The UK rail dataset shows much denser clusters of stores than the 33 kV substations and US airports. Therefore, to be able to model a range of real world networks an algorithm must be able to generate synthetic networks with a different number, location and density of clusters of nodes. The algorithm must also be able to generate individual clusters of nodes where the radius of the cluster increases rapidly with the addition of the first few nodes and then only increases slightly with the addition of further nodes.

Table 4
Showing the input values used to generate the proxy nodal layouts.

Dataset	EATN	CATN
Proportion of Seed Nodes	~4%	~4%
Range of starting radii used	800, 900, 1000	800, 1000
C_D (global density)	50	100
Proportion of Nodes allowed to form outside the influence of a cluster	40%	30%

3. Development of the clustering algorithm

In a similar manner to cellular automata, our proposed algorithm requires the input of a set of initial conditions, from which the nodal layout forms over a given timeframe. These initial conditions define the spatial boundary of the network, the number of starting or seed nodes (which form a small proportion of the total number of nodes in the network and also form the centre of each cluster) and the location and initial radius of these seed nodes. A summary of the input parameters is provided in Table 1 and discussed in further detail in this section (with a detailed application provided in Section 4). Using these inputs the network is allowed to grow as the remaining nodes are added individually to the network at each time-step until the total number of nodes is reached (e.g. simulating the opening of new airports and rail stations for example). In this approach, we acknowledge that decision over where to place an actual railway station, or airport, is not made based upon 'simple' rules, but upon complex regulations and social-economic issues; however, we argue that this is not significant as we are not aiming to replicate the actual networks exactly, but rather form synthetic configurations that contain the same aggregate features and can therefore be used in planning studies as alternative futures or for generic hazard tolerance assessments.

At each time-step the algorithm determines if an added node will be located within the radius of one of the individual clusters or will be located outside the influence of all of the clusters, depending on a user specified probability. By allowing a small proportion of the total number of nodes in the network to be located outside the cluster radii, a rural environment over the whole of the spatial domain is represented. However, if the added node is not to be located in a rural area, then this node is 'attracted' to one of the individual clusters. This strength of this attraction is determined using a 'pulling-power' factor whose value is dependent upon the density of the cluster and is calculated using Eq. (1). The pulling-power is not fixed for the whole analysis but rather is recalculated after each node is added. The pulling-power encompasses the idea that a city,

Table 3
Showing the Actual Nearest Neighbour value for the datasets and also the mean, max/min values for the 10 generated proxy layouts.

Dataset	Actual Layout	Generated Layout	
	Average Nearest Neighbour Value	Mean Average Nearest Neighbour Value	Min/Max Average Nearest Neighbour Value
UK rail stations	0.46	0.44	0.42/0.48
US airports	0.76	0.78	0.76/0.79
33 kV Substations	0.75	0.75	0.72/0.79

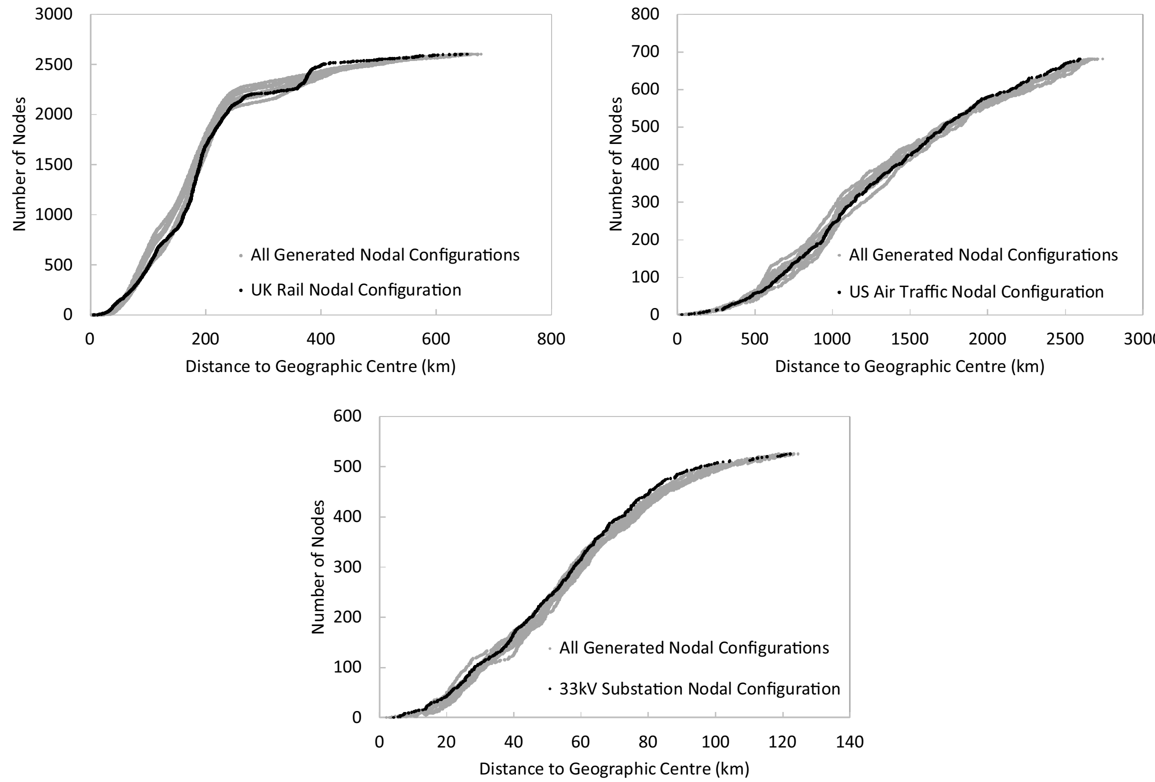

Fig. 5. Showing all 10 generated spatial distributions (grey) compared to the actual distribution of nodes (black) for (a) UK rail stations, (b) US airports and (c) 33 kV substations.

with a high population density, can be expected to have more nodes (representing railway stations, for example) than a rural community which has a significantly lower population density.

$$Pp = \frac{number\ of\ nodes_{CLUSTER}}{radius_{CLUSTER}} \tag{1}$$

With the addition of a new node to the cluster, the radius of the cluster is allowed to expand outwards, in order to simulate the 'growth' pattern of the individual clusters in the network as shown by the real world networks (see Fig. 2). A logarithmic trend line has been fitted to the individual clusters, shown in Fig. 2, and it can be seen that this is a reasonable approximation for the real world data. The expansion of the radius, following the addition of a new node, is controlled using Eq. (2). It is worth noting that this equation is not used to calculate the original radius assigned to the cluster, but only to calculate the subsequent growth.

$$Radius = C_D\,(\ln(number\ of\ nodes) + 1) \tag{2}$$

The C_D term in Eq. (2) controls the average density of the whole nodal layout (i.e. the global density) and its effects can be seen in Fig. 3. In this figure two clustered layouts have been generated using the same initial inputs (i.e. seed location, initial radius size) but a different C_D value.

The density of each individual cluster, relative to the other clusters, is determined by changing the initial radius assigned to each seed node; assigning a large radius results in a low density cluster, whilst a small radius results in a dense cluster (an example of this is shown in Fig. 4).

4. Initial assessment of clustering algorithm

To initially assess the ability of the algorithm to generate synthetic proxies for real world nodal configurations, we generate layouts for the three real world networks shown in Fig. 1, our "development datasets". We then validate the algorithm by using it to general proxies for two real world spatial networks that have not been used to inform the algorithm. In both cases, we compare the properties of these datasets by plotting the spatial distribution (as shown in Fig. 1(b, d)) and also compare the Average Nearest Neighbour value of each layout (Ebdon 1977; ArcGIS, 2013).

To generate proxies for the US airport dataset, we firstly divide the land mass of the USA into a 20 km grid; whilst to generate the UK rail network we divide the land mass of the UK into a 5 km grid and for the 33 kV substation dataset we divide the UK into a 1 km grid. This gradation is required to ensure nodes only form over a land mass (i.e. not over the ocean or other bodies of water) and its resolution is based on a compromise between accuracy and computational expense. For all three real world spatial networks we generate 10 synthetic proxies, which are generated using the set of initial conditions shown in Table 2. The locations and sizes of the starting radii, in all layouts, were identified by viewing population density maps; seed nodes were located over areas of highest population density and their radius value assigned based upon the density of the population. We used either two, or three, values of radii to represent population areas with different densities in these simulations, representing town, city and village, depending on the most valid for the dataset; however, we could have assigned each cluster its own radii value based upon its population. The C_D value used in the simulations was determined by analysing a number of clusters in each real world network. Individual clusters were isolated, using the Kernel Density images, and then their radius and number of nodes in the cluster were identified and Eq. (2) used to calculate the C_D value for the cluster. The choice to allow a proportion of nodes to form outside the influence of clusters was based upon the global 'spread' of actual stores, or stations, over the study area (informed by assessing the Kernel density images and also the Average Nearest Neighbour values). For example, if the clusters are tight then only a very small proportion of nodes were allowed to

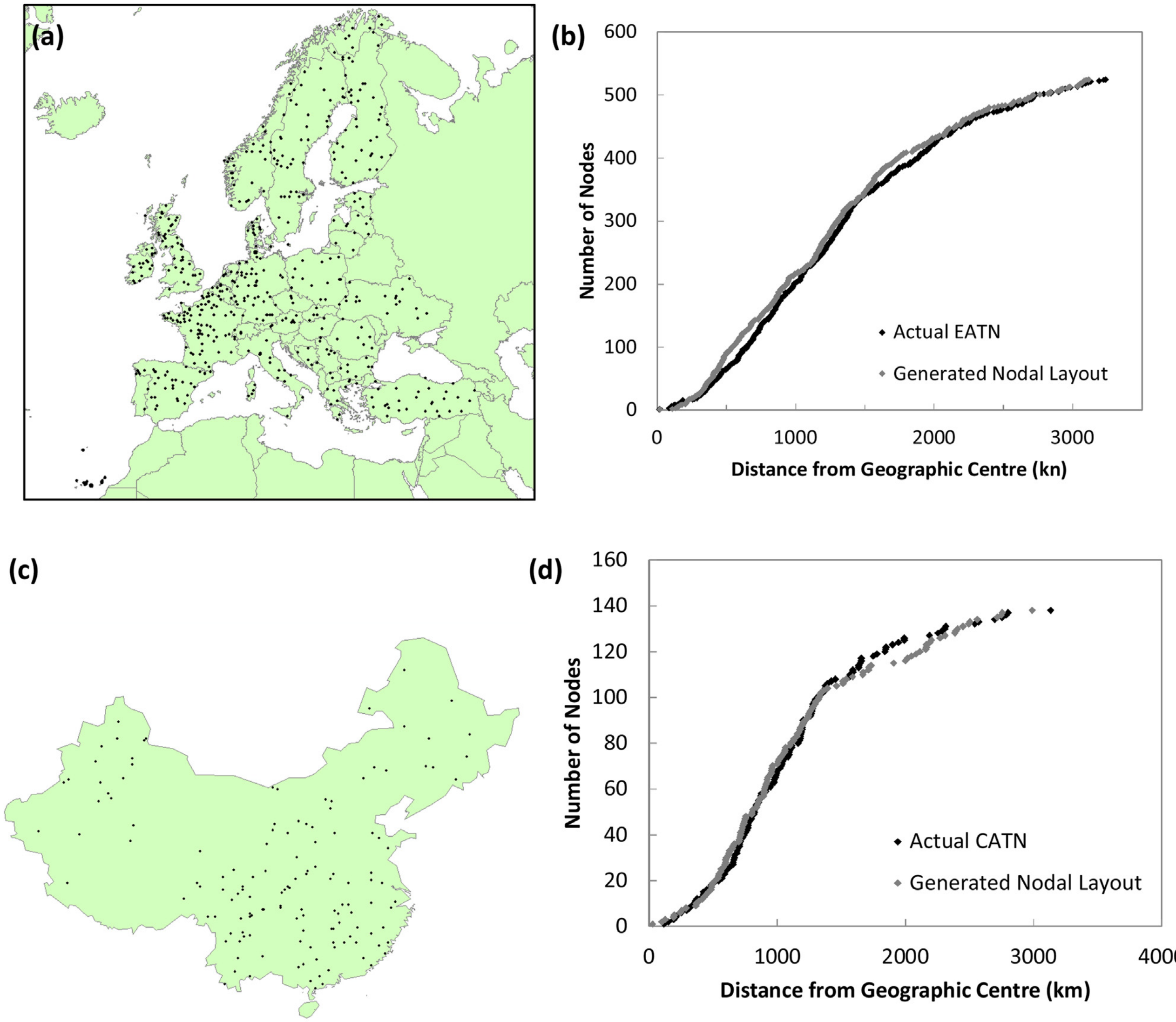

Fig. 6. The generated nodal layout for the (a) European air traffic network and (c) China air traffic network, where the black dots represent the airport locations and a comparison for the spatial distribution of nodes for the (b) European air traffic network and (d) China air traffic network (black) and the generated nodal layout shown in (a, c) (grey).

form outside the influence of clusters. It is worth noting that we do not change any input values to generate the 10 synthetic nodal layouts (shown in Fig. 5).

The resulting spatial distributions for these proxy nodal configurations have been shown, and compared to the real world configurations, in Fig. 6. In this figure, we have chosen to plot all 10 synthetic nodal configurations on one graph, to highlight the small spread of data. From this figure, it can be seen that for all three real world spatial layouts, the synthetic layouts have very similar spatial distributions.

We also validate the algorithm, by calculating the Average Nearest Neighbour value and compare this to the value for the actual datasets, for 10 generated layouts (Table 3). The Average Nearest Neighbour value is a measure of how clustered, or dispersed, a spatial layout is and is based on the average distance from each node to its nearest node (Ebdon 1977). From Table 3, it can be seen that the generated nodal layouts are a good proxy for the real world layouts, with similar values of mean Average Nearest Neighbour and also only a small spread in the results achieved (when viewing the maximum and minimum values).

5. Validation of clustering algorithm and combination with traditional network generation algorithms

In this paper, we have so far generated synthetic nodal distributions for the US airport, UK rail and a section of UK 33 kV electrical substations to show that the clustering algorithm can generate nodal distributions with different characteristics that have similar properties to their real-world counterparts. We now validate the algorithm, by using it to generate proxy nodal layouts for the airports of the EATN (which contains 525 airports) and the China air traffic network (CATN) (containing 138 airports), our two "validation datasets". These two nodal layouts were not used to inform the algorithm, or used it its development, and are used only to validate the algorithm. Once we have generated the nodal layouts we show how these can be coupled with a traditional network generation algorithm, developed by Wilkinson et al. (2011) (which provides a 'rule' set determining the formation of links) to form a synthetic network, which has both the same spatial and topological properties as the real networks.

To generate proxy nodal layouts for the EATN and the CATN we again grid the land mass of the study area (i.e. European and China

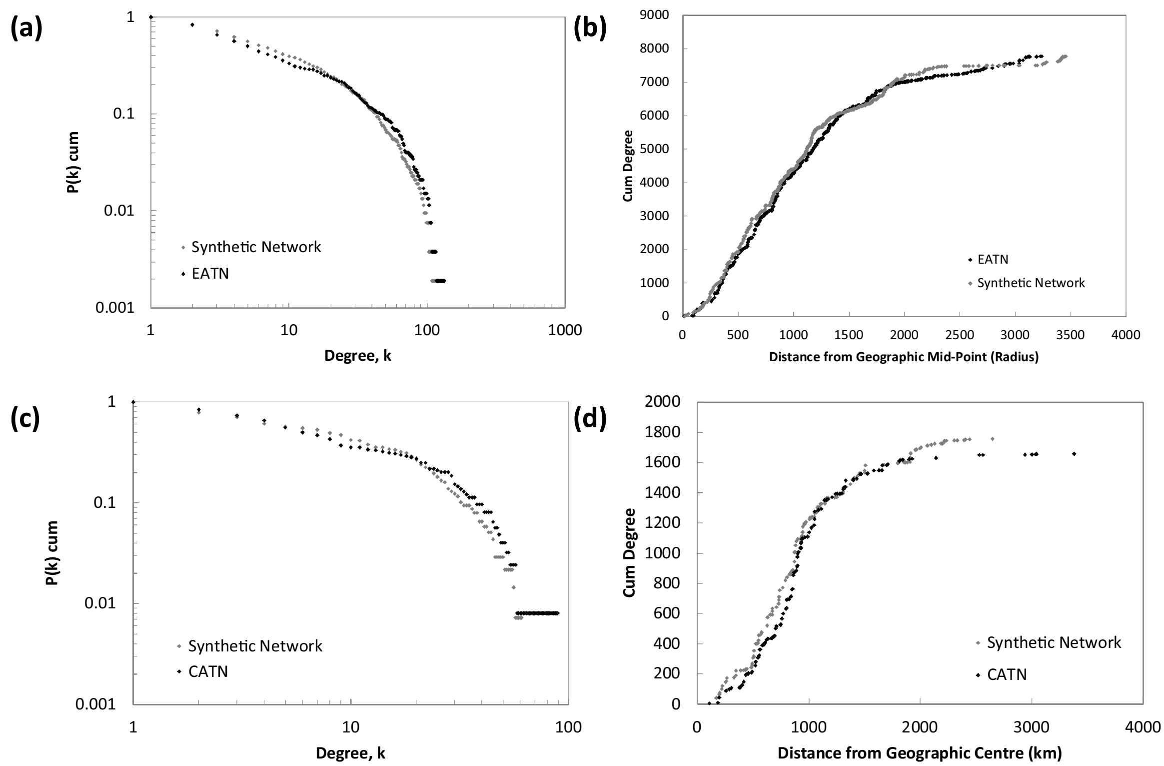

Fig. 7. Showing the (a, c) degree distributions and (b, d) spatial degree distributions for the European air traffic network and China air traffic network, respectively. In all graphs, the actual networks are shown in black and a generated proxy networks, generated using the same clustered nodal layout as shown in Fig. 6, are shown in grey.

airspace) into a 20 km grid and only allow nodes to form in the grid cells. The parameters used to generate both nodal configurations are shown in Table 4, whilst the generated nodal layouts are shown in Fig. 6(a, c) and the spatial distributions are shown in Fig. 6(b, d). Comparing the spatial distribution to that of the actual EATN and CATN shows that the generated nodal layouts are a good proxy for that of the real world networks.

In their paper, Wilkinson et al. (2011) developed an algorithm to generate synthetic networks for the EATN; however this algorithm did not reproduce the small scale features such as clustering of airports as it made the simplifying assumption that the nodes were distributed uniformly with distance. Therefore, we now combine this network generation algorithm with our nodal configuration to form a proxy network for the EATN and CATN, where both the 'rules' governing the formation of nodes and links have been considered. As this network generation algorithm incorporates the idea of *growth* the order in which nodes are added to the network will influence the placement of the higher degree nodes (as nodes that are introduced early to the network have more chances to 'attract' links from new nodes). We include in our simulation a similar geo-political constraint to that of Guimera and Amaral (Guimera & Amaral 2004) in that we have added one node to each country, or province, in order of population (highest to lowest) and have then added the remainder of nodes randomly to the network. This procedure is necessary to simulate the dispersion of high degree airports observed in the EATN and CATN. The degree distribution and spatial degree distribution for these synthetic networks have been shown in Fig. 7, where they are compared to those of the actual air traffic networks. It can be seen from this figure, that both of these distributions are in good agreement with the actual air traffic networks; therefore, it can be concluded that it is possible to generate a proxy network where both the 'rules' governing the locations of nodes and links have been considered.

6. Conclusions

In this paper, we have presented an algorithm to generate proxy nodal layouts for real world networks as well as more general nodal layouts. The algorithm is based upon, but simplifies, the rules of cellular automata and only requires recent population density information for the study area. From this data the initial conditions, regarding the location and radius of the seed nodes, can be determined. We have used this algorithm to generate proxy nodal layouts for five real world networks, which have the same spatial distribution (both local and global) of nodes and growth pattern of individual clusters as their real world counterparts. Our validation has demonstrated that the growth rate for individual clusters is exponential and that clusters of different densities can be formed by assigning seed nodes different radii values.

To show that these nodal layouts can be combined with traditional network generation algorithms (where only the formation of links is considered important), to form fully synthetic spatial networks, we have combined the proxy nodal layouts of the EATN and the CATN with the generation algorithm of Wilkinson et al. (2011) to form two proxy networks where both the 'rules' behind the location of nodes and formation of links has been considered. The importance of these algorithms is that they can be used to assess generic features (such as resilience or efficiency) likely to be common to a range of networks.

Perhaps more importantly, as the rule sets are mainly based on population information, the model can also be used to generate alternative futures in planning exercises, for example to explore the impact of policy, population changes (or shifts) and increasing urbanisation (in a similar manner to Fu, Wilkinson, and Dawson (2016). The resulting spatial networks can then be assessed to understand their characteristics and resilience to hazard. These 'future scenarios' can be readily incorporated into the model through the alteration of the input parameters. For example, decreasing the proportion of nodes allowed to form outside the influence of individual clusters wold simulate an increased urbanised environment and the impact of different population densities can be explored by changing the starting radii of the seed nodes. The impact of "new" cities (e.g. garden cities) can be incorporated into the model through the addition of a new seed node at a later timestep in the model. We argue that this is an important area for future research and characterising the effects that these parameters have to the formation of real world systems are likely to lead to an increased understanding of the effect of population, political and social influences.

Acknowledgement

This research was partly funded by the Engineering and Physical Sciences Research Council, UK, in the form of an EPSRC Doctoral Prize Fellowship (EP/I035781/1) awarded to S. Dunn and their support is gratefully acknowledged.

References

Albert, R., Jeong, H., & Barabasi, A. L. (1999). Internet—diameter of the world-wide web. *Nature*, *401*(6749), 130–131.
Albert, R., Jeong, H., & Barabasi, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378–382.
ArcGIS. (2013). *Average nearest neighbor (spatial statistics)*.. Retrieved 17 November 2014, from. http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000008000000
Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.
Batty, M. (2008). The size, scale, and shape of cities. *Science*, *319*, 769–771.
Batty, M. (2013). A theory of city size. *Science*, *340*, 1418–1419.
Bettencourt, L. M. A. (2013). The origins of scaling in cities. *Science*, *340*, 1438–1441.
Bettencourt, L. M. A., Lobo, J., Helbing, D., Kuhnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(17), 7301–7306.
Bettencourt, L. M. A., Lobo, J., & West, G. B. (2008). Why are large cities faster? Universal scaling and self-similarity in urban organization and dynamics. *European Physical Journal B*, *63*(3), 285–293.
Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics Reports-Review Section of Physics Letters*, *424*(4-5), 175–308.
Carvalho, R., Buzna, L., Bono, F., Gutierrez, E., Just, W., & Arrowsmith, D. (2009). Robustness of trans-European gas networks. *Physical Review E*, *80*(1).
Clarke, K. C., & Gaydos, L. (1998). Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, *12*(7), 699–714.
Clarke, K. C., Hoppen, S., & Gaydos, L. (1995). A self-modifying cellular automation model of historical urbanisation in the San Francisco Bay area. *Environment and Planning B-Planning & Design*, *24*, 27–261.
Crucitti, P., Latora, V., & Marchiori, M. (2004). A topological analysis of the Italian electric power grid. *Physica a-Statistical Mechanics and Its Applications*, *338*(1-2), 92–97.
data.gov.uk. (2015). *Network rail—inspire data*.. Retrieved 5th October 2015, from. https://data.gov.uk/dataset/railway-network-inspire
Dunn, S., Fu, G., Wilkinson, S., & Dawson, R. (2013). Network theory for infrastructure systems modelling. *Proceedings of the ICE—Engineering Sustainability*, *166*(5), 281–292.
Ebdon, D. (1977). *Statistics in geography*. Oxford: Blackwell Publishers.
Erdos, P., & Renyi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institutre of the Hungarian Academy of Sciences*, *5*, 17–61.
Florida, R. (2011). *Why cities matter*. Retrieved 10 February 2016, from. http://www.citylab.com/design/2011/09/why-cities-matter/123/
Fu, G., Wilkinson, S., & Dawson, R. (2016). A spatial network model for infrastructure system development. *Computer-Aided Civil and Infrastructure Engineering*, *31*, 661–680.
Gastner, M. T., & Newman, M. E. J. (2006a). Optimal design of spatial distribution networks. *Physical Review E*, *74*(1).
Gastner, M. T., & Newman, M. E. J. (2006b). The spatial structure of networks. *European Physical Journal B*, *49*(2), 247–252.
Gorman, S. P., & Kulkarni, R. (2004). Spatial small worlds: New geographic patterns for an information economy. *Environment and Planning B-Planning & Design*, *31*(2), 273–296.
Guimera, R., & Amaral, L. A. N. (2004). Modeling the world-wide airport network. *European Physical Journal B*, *38*(2), 381–385.
Knight, Frank (2015). The Wealth Report.
Kwan, S. C., & Hashim, J. H. (2016). A review on co-benefits of mass public transportation in climate change mitigation. *Sustainable Cities and Society*, *22*, 11–18.
Liu, J. Z., & Tang, Y. F. (2005). An exponential distribution network. *Chinese Physics*, *14*, 643–645.
Lizarralde, G., Chmutina, K., Bosher, L., & Dainty, A. (2015). Sustainability and resilience in the built environment: The challenges of establishing a turquoise agenda in the UK. *Sustainable Cities and Society*, *15*, 96–104.
Lordan, O., Sallan, J. M., & Simo, P. (2014). Study of the topology and robustness of airline route networks from the complex network approach: A survey and research agenda. *Journal of Transport Geography*, *37*, 112–120.
Miller, W. (2015). What does built environment research have to do with risk mitigation, resilience and disaster recovery? *Sustainable Cities and Society*, *19*, 91–97.
Openflights. (2010). *OpenFlights.org*. Retrieved 13th August 2010, from. http://openflights.org/
Ruza, J., Kim, J. I., Leung, I., Kam, C., & Ng, S. Y. M. (2015). Sustainable, age-friendly cities: An evaluation framework and case study application on Palo Alto, California. *Sustainable Cities and Society*, *14*, 390–396.
Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., & Manna, S. S. (2003). Small-world properties of the Indian railway network. *Physical Review E*, *67*(3).
Sharifi, A. (2016). From Garden City to Eco-urbanism: The quest for sustainable neighborhood development. *Sustainable Cities and Society*, *20*, 1–16.
Sole, R. V., Rosas-Casals, M., Corominas-Murtra, B., & Valverde, S. (2008). Robustness of the European power grids under intentional attack. *Physical Review E*, *77*(2).
Spataru, C., Drummond, P., Zafeiratou, E., & Barrett, M. (2015). Long-term scenarios for reaching climate targets and energy security in UK. *Sustainable Cities and Society*, *17*, 95–109.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442.
Wilkinson, S., Dunn, S., & Ma, S. (2011). The vulnerability of the European air traffic network to spatial hazards. *Natural Hazards*, *60*(3), 1027–1036.
Wu, F. (2002). Calibration of stochastic cellular automata: The application to rural-urban land conversions. *International Journal of Geographical Information Science*, *16*(8), 795–818.
Zhang, X., Miller-Hooks, E., & Denny, K. (2015). Assessing the role of network topology in transportation network resilience. *Journal of Transport Geography*, *46*, 35–45.